**Acute sport concussion assessment optimization:**
**A prospective assessment from the CARE Consortium**

Steven P. Broglio PhD
Michigan Concussion Center
University of Michigan

Jaroslaw Harezlak, PhD
Department of Epidemiology and Biostatistics
Indiana University, Bloomington, IN

Barry Katz, PhD
Department of Biostatistics
Indiana University School of Medicine

Shi Zhao, PhD
Eli Lilly and Co.
Indianapolis, IN

Thomas McAllister, MD
Department of Psychiatry
Indiana University

Michael McCrea, PhD ABPP
Department of Neurosurgery
Medical College of Wisconsin

and the CARE Consortium Investigators

**Corresponding Author**
Steven Broglio, PhD
401 Washtenaw Ave
Ann Arbor, MI 48109
broglio@umich.edu

This is the author's manuscript of the article published in final edited form as:
Broglio, S. P., Harezlak, J., Katz, B., Zhao, S., McAllister, T., McCrea, M., Hazzard, J., Kelly, L., Campbell, D., Jackson, J., McGinty, G., O'Donnell, P., Cameron, K., Susmarski, A., Goldman, J., Giza, C., Buckley, T., Kaminski, T., Clugston, J., … CARE Consortium Investigators. (2019). Acute Sport Concussion Assessment Optimization: A Prospective Assessment from the CARE Consortium. Sports Medicine, 49(12), 1977–1987. https://doi.org/10.1007/s40279-019-01155-0

**Acknowledgements:**
The authors would like to thank April Reed Hoy (Azusa Pacific University), Justus Ortega (Humboldt State University), Nicholas Port (Indiana University), Margot Putukian (Princeton University), Dianne Langford and Ryan Tierney (Temple University), Holly Benjamin (University of Chicago), Kevin Guskiewicz and Jason Mihalik (University of North Carolina–Chapel Hill), Jessica Miles (University of North Georgia), Jeffrey Bazarian (University of Rochester), Stefan Duma (Virginia Tech), Christopher Bullers and Christopher Miles (Wake Forest University), Brian Dykhuizen (Wilmington College), Laura Lintner (Winston-Salem State University), Paul Pasquina (Uniformed Services University), Jody Harland, Janetta Matesan, Larry Riggen (Indiana University), Ashley Rettmann (University of Michigan), Melissa Koschnitzke (Medical College of Wisconsin), Michael Jarrett, Vibeke Brinck, and Bianca Byrne (Quesgen), Thomas Dompier, Christy Collins, Melissa Niceley Baker, and Sara Dalton (Datalys Center for Sports Injury Research and Prevention), and the research and medical staff at each of the participating sites.

**ABSTRACT**

**Background:** Numerous medical organizations recommend a multifaceted approach to the assessment of concussion occurring during sporting events. A number of tools are available to clinicians with a wide breadth of sensitivity and specificity. Little work has been done to evaluate the combined efficiency of these tools in concussed male and female athletes from a broad array of collegiate sports and with variable time from the pre-season baseline evaluation.

**Objective**: To optimize the concussion assessment battery for application within the first 72 hours of injury and identify the necessary baseline retesting frequency.

**Methods**: Between 2014 and 2017, a total of 1,458 NCAA athletes sustaining 1,640 diagnosed concussions completed a baseline assessment each year of the investigation and were evaluated up to three times within the first 72 hours of injury using a standardized assessment protocol. Classification and regression tree analyses were implemented to identify the most efficient multifaceted assessment pathway to quantify concussion-related outcomes. Results were optimized for assessments occurring within one hour post-injury, 1 to 24 hours post-injury, and 24 to 72 hours post-injury when using the raw post-injury assessment performance, difference scores from baseline evaluations occurring in the same year, and difference scores from baseline evaluations occurring the year prior.

**Results**: At each of the assessment time points, the analyses indicated that alone or in combination, a symptom evaluation, BESS scores collected on the firm surface, and SAC total score offered the best overall performance when compared to pre-morbid performance captured in the same season. Optimized sensitivity of the multifaceted approach was 61% within 1 hour of injury, 67% at the 1 to 24 hour interval, and 55% at the 24 to 72 hour interval when difference scores from the same season baseline were available.

**Conclusions**: This investigation identified key concussion assessments in quantifying post-concussion performance among student athletes that were maximized when same season pre-morbid evaluation

were available. Consistent with clinical recommendations, medical professionals should continue to focus on symptom reporting, postural control, and neurocognitive screening to support the clinical examination when making a concussion diagnosis.

**KEY POINTS**

- Within 72 hours of concussion, symptoms, BESS (firm surface), and the SAC offer the greatest information supporting the clinical examination
- Post-injury testing was optimized when compared to a same season baseline
- Neurocognitive testing both within and beyond the 72 hour mark may be beneficial in specific circumstances.

## 1. INTRODUCTION

Despite decades of research, diagnosis of sport-related concussion (SRC) continues to be based on subjective symptom-reporting and the clinical examination[1]. A number of objective measures have been developed for sideline and clinical use, but none have demonstrated psychometrics at the necessary level to replace the clinical exam as the diagnostic standard[2, 3]. Logistical and practical concerns also influence how these tools are implemented clinically. For example, a neurocognitive screening tool demonstrates high sensitivity immediately following injury, but fails to demonstrate these same levels 72 hours later[2]. A similar decline in sensitivity has been noted with a semi-objective assessment of balance[2]. Neurocognitive testing, once touted as the 'cornerstone' of concussion assessment, has demonstrated variable reliability and sensitivity[4, 5], limiting its utility in isolation of other measures. The best and most broadly implemented assessment is the use of athlete-reported symptoms, which demonstrates sensitivity levels as high as 0.89 immediately following injury[2]. Use of symptoms, however, is limited by an individual athlete's awareness of concussion symptoms and their willingness to be forthright in reporting symptoms, which may vary by sport and gender[6].

The limitations of these measures when used in isolation has driven several organizations to recommend a multi-dimensional approach to concussion assessment, whereby a combination of assessments evaluating neurocognitive functioning, motor control (i.e., balance) and post-concussive symptoms are used to support the clinical examination[1, 7-9]. In this model, changes noted on any single test outside of normal variance could indicate concussion-related impairment. The combined assessment approach has proven beneficial with sensitivities improving to 0.94 immediately post-injury when neurocognitive screening, balance, and symptoms are implemented[2] and with similar findings when a robust neurocognitive assessment, computer-instrumented balance, and athlete-reported symptoms are administered within 24 hours of injury[3].

Intertwined with how to best apply clinical assessments during post-concussion evaluations is the utility of the baseline assessment. The baseline assessment provides a snapshot of brain functioning that is typically captured prior to the competitive season. Although some suggest that baseline testing is not necessary because of cost and insufficient benefit in the post-injury assessment[10, 11], some organizations continue to recommend their use[7, 9]. This stance is particularly true for those participating in contact and collision sports, but how often the baseline assessment should be given during an academic career is a matter of debate among collegiate student-athletes. Annual assessments have been suggested for interscholastic athletes who are undergoing rapid brain growth and development[12], but there is no consensus surrounding how frequently collegiate athletes should be evaluated (i.e. annually, every other year, etc.). Finding an evidence-based answer to this question has significant resource and monetary implications for the medical staff administering the tests and time costs to the athletes completing the examinations.

As such, the primary purpose of this study was to evaluate the sensitivity of commonly implemented concussion assessment tools immediately following injury and up to 72 hours post-injury in a nationally representative sample of concussed male and female athletes from a broad array of collegiate sports. The secondary purpose was to establish the necessary frequency of the baseline assessment by evaluating sensitivity of the same measures when the baseline assessment occurred one and two years prior.

## 2. METHODS

As part of a larger investigation on the natural history of concussion, all male and female varsity-level athletes from 29 NCAA member institutions across the United States were approached to enroll in the

Concussion Assessment, Research, and Education (CARE) Consortium. Non-varsity sport cadets at the military service academies were also approached for enrollment. The CARE research design and methods have been described in detail elsewhere[13]. In brief, following written informed consent, all participants completed a pre-season baseline assessment that included an in-depth demographics and medical history questionnaire, measures of neurocognitive functioning (note: only one type of neurocognitive assessment per site), balance, and self-reported symptoms. Athletes from a subset of participating schools completed self-selected emerging assessments. Table 1 outlines the number of sites contributing data for each metric. Athletes participated in their sport without interference from the investigative team and were monitored for concussion by the local medical staff. When a concussion was suspected, the athlete completed a measure of neurocognitive function, balance assessment, and symptom report within 6 hours of injury, 24-48 hours post-injury, when cleared to begin the return to play process, and when cleared for unrestricted return to play. An evidence-based review was used to define concussion across the participating sites[14]. Emerging assessments were re-administered at these same time points. The baseline assessment required 55-60 minutes to complete, while the 6-hour time point took approximately 20 minutes and assessments at the remaining time points took 35-40 minutes. Each CARE performance site obtained local Institutional Review Board and US Army Human Research Protection Office approval prior to data collection. This study was completed in accordance with the Declaration of Helsinki.

Insert Table 1 here

Between the study start in 2014 and the end of the Fall 2017 season, 34,634 varsity sport athletes and non-varsity sport cadets had been enrolled and completed at least one baseline evaluation. To evaluate the post-concussion sensitivity of assessments identified above, 10,048 non-varsity sport cadets were removed from the dataset to be evaluated in a forthcoming analysis. Among the 24,586 remaining varsity sport athletes, 2,543 total concussions were captured, including 474 concussions that occurred outside of sport (e.g., motor vehicle accident) which were removed from the analyses.

*2.1 Data Analysis*

As data were not always captured exactly within the post-injury windows identified above, intervals of 0-1.25hrs (sideline), 1.25-24hrs (post-event), and 24-72hrs (clinic) were established using the day and time of the post-injury assessments to capture clinically relevant time points. Data were analyzed based on: (a) the raw post-concussion scores for the concussed athletes and baseline scores for the non-concussed controls; (b) within subject difference between the post-concussion scores and same season baseline for the concussed athletes and differences between 2 baseline scores for the non-concussed controls; and (c) difference between the post-concussion scores and the prior season baseline for the concussed athletes and differences between 2 baseline scores for the non-concussed controls. The change score relative to the same season baseline and/or baseline one year prior were first compared to previously derived test-retest confidence intervals (i.e., 75%, 85%, 90%, 92.5%, 95%, 97.5%, 99%) from the CARE Consortium[15]. A test was deemed to be sensitive at a given confidence interval if it met or exceeded the change score (i.e., worse performance) for that level. For example, a 9-point increase in BESS test performance would elicit 95% likelihood the increase was not due to chance.

In addition to establishing the sensitivity of each individual assessment, classification and regression trees (CART) methodology was utilized to identify the best combination of factors from a multifaceted concussion assessment battery. Classification trees are a nonparametric way of assessing the influence of multiple factors, simultaneously allowing for multi-way interactions of the measures. CART analysis allows the derivation of a small number of factors predictive of concussions along with their interactions from a number of candidate factors, without being limited by the highly correlative nature of some of the measures, or their nonlinear effects and multi-way interactions. At each level of the analysis, all available concussion assessments are examined and the measure that maximizes group homogeneity (i.e., concussed or non-concussed) selected. The process repeats at the next level with the remaining assessments.

In this approach, we entered scores from the mandatory assessments (total SCAT symptom score, SCAT symptom severity score, SAC total score, BESS total score, BESS firm only score, BESS foam only score, BSI-18 and ImPACT sub-test scores (verbal memory, visual memory, visual motor speed, and reaction time)) simultaneously in the CART analysis. Emerging tests were not included due to limited data availability that would have generated unstable models (i.e. <100 concussions with all mandatory and at least one emerging assessment). The resulting tree includes the optimal factor combinations producing the final classification of individuals with and without concussions. The performance of the CART approach is assessed directly from the misclassification error ("concussed" classified as "non-concussed" and vice versa) and measures of sensitivity, specificity, and positive and negative predictive value estimates. Analyses were performed using different weights assigned to misclassification errors as the number of concussed athletes was much smaller than the number of non-concussed athletes included in the analyses. To compare trees and select the best classification pathway, we maximized an F1 score which combines the positive predictive value (PPV) and sensitivity. Specifically, F1 score is calculated as a harmonic mean of the PPV and sensitivity. We chose F1 score as the criterion, since we emphasize the correct classification of concussions and put less emphasis on true negatives (i.e., correct classification of non-concussed athletes).

## 3. RESULTS

Adequate data from a maximum of 1,640 incident concussions, occurring in 1,458 athletes (63.0%male, 19.0±1.2 years, 179.7±11.5 cm, 84.1±21.7kg, 0.60±0.85 self-reported previous diagnosed and undiagnosed concussions) were included for analyses. Among the included concussions, 1,640 had a baseline assessment that coincided with the same academic year (124.1±88.0 days prior to injury), while 770 also had a baseline assessment from the year prior (475.9±83.8 days prior to injury).

3.1 Individual Assessment Performance

*3.1.1 Sideline Evaluation*

Sensitivity statistics for the mandatory measures across a range of confidence intervals are presented in Supplemental Table 1. For those with a same season baseline, concussion-related outcomes were best quantified using the SCAT symptom evaluation, which had the highest sensitivity across the confidence intervals. This finding held for both the total number of symptoms (17-84%) reported by the athlete and symptom severity (25-88%), although symptom severity performed better at the higher confidence intervals. The SAC (10-55%) and BESS (4-46%) were less sensitive across the same confidence intervals. For those with a baseline occurring one year prior, the SCAT symptom evaluation again showed the highest sensitivity across all confidence intervals. This held for both the total number of symptoms (13-77%) reported by the athlete and symptom severity score (18-78%), with symptom severity performing better at the higher confidence intervals. The SAC (12-57%) and BESS (1-40%) were less sensitive across the same confidence intervals. There was insufficient data to establish sensitivity of the computer-based neurocognitive assessments (Supplemental Table 2.) and the BSI-18 (Supplemental Table 1.) at this post-injury window when baselines were completed in the same season or one year prior.

Among the emerging assessments, subscores of the VOMS yielded the highest sensitivities for those with a baseline assessment in the same season (Supplemental Table 3). This included the Smooth

Pursuit (55-100%), Horizontal Saccades (54-87%), Vertical Saccades (63-89%), Near Point Convergence (NPC) Symptoms (43-88%), NPC (9-64%), Horizontal VOR (47-88%), Vertical VOR (45-86%), and Visual Motion Sensitivity (VMS) (55-86%). Sensitivity of the K-D ranged from 21-39% and there was an insufficient number of concussions to evaluate the RTclin (Supplemental Table 3.). When baselines captured one year prior were considered, the VOMS again yielded the highest results: Smooth Pursuit (63-100%), Horizontal Saccades (58-79%), Vertical Saccades (63-84%), NPC Symptoms (37-79%), NPC(0-42%), Horizontal VOR (56-83%), Vertical VOR (56-89%), and VMS (61-83%). Sensitivity of the K-D ranged from 15-54% and not enough RTclin data were available for the analyses.

*3.1.2 Post-event Evaluation*

The evaluation of the 1 to 24 hours post-injury sensitivity of the mandatory measures for those with a same season baseline indicated that the athlete symptom reports were the highest across the confidence intervals (Supplemental Table 1.). Symptom severity (25-81%) out performed symptom total (21-82%) only at the highest confidence interval. Performance on the SAC (7-54%) followed as the next highest, and then the BSI-18 (2-47%) and BESS (3-35%). Computer-based neurocognitive testing using the ImPACT test showed consistency across the test sub-scores: Verbal Memory (3-37%), Visual Memory (3-36%), Visual Motor Speed (3-31%), and Reaction Time (3-37%). There were insufficient data to evaluate the CCAT and CNS Vital Signs at this time point. For those with a baseline examination occurring one year prior, athlete-reported symptoms again yielded the highest sensitivities with similar performance between symptom total (21-80%) and symptom severity (22-79%). The SAC (8-52%), BSI-18 (3-52%), and BESS (4-38%) were all slightly less sensitive. Sensitivity of the ImPACT subscores using baseline performance from one year prior were: Verbal Memory (2-31%), Visual Memory (2-33%), Visual Motor Speed (5-31%), and Reaction Time (3-37%). There were insufficient data to evaluate the CCAT and CNS Vital Signs (Supplemental Table 2).

Performance on the emerging assessments (Supplemental Table 3) indicated the subscores of the VOMS yielded the highest sensitivities for those with a baseline assessment in the same season. This included the Smooth Pursuit (48-100%), Horizontal Saccades (44-83%), Vertical Saccades (46-82%), NPC Symptoms (38-80%), NPC (6-47%), Horizontal VOR (39-81%), Vertical VOR (38-81%), and VMS (43-81%). Sensitivity of the K-D (16-48%), and RTclin(0-27%) followed. For those with a baseline examination occurring one year prior, VOMS sensitivities were similar for Smooth Pursuit (40-100%), Horizontal Saccades (39-88%), Vertical Saccades (42-87%), NPC Symptoms (35-84%), NPC(4-46%), Horizontal VOR (38-84%), Vertical VOR (36-84%), and VMS (37-82%). Sensitivity of the K-D was 5-41% under the same parameters and not enough data were available to evaluate the RTclin.

*3.1.3 Clinic Evaluation*

Performance among the mandatory measures 24-72 hours post-injury (Supplemental Table 1.) indicated that symptom total (17-70%) and symptom severity (17-66%) were the most sensitive to lingering effects of injury for those with a same-season baseline. This was followed by the BSI-18 (3-51%), SAC (4-47%), and BESS (2-30%). Computer based neurocognitive testing (Supplemental Table 2.) indicated similar performance between the ImPACT (Verbal Memory (2-34%), Visual Memory (2-32%), Visual Motor Speed (4-32%), Reaction Time (3-38%)), the CCAT (Processing Speed (8-38%), Attention (7-36%), Learning (0-15%), Working Memory Speed – Speed (0-15%)), and CNS Vital Signs (Neurocognition (0-57%), Composite Memory (0-39%), Verbal Memory (0-29%), Visual Memory (0-18%), Psychomotor Speed (2-28%), Reaction Time (10-45%), Complex Attention (2-50%), Cognitive Flexibility (2-48%), Processing Speed (0-26%), Executive Function (2-42%), Simple Attention (8-34%), and Motor Speed(4-38%). When compared to baseline scores captured the year prior to injury, symptom total (15-68%) and symptom severity (13-65%) yielded the highest sensitivity, followed by the BSI-18 (2-50%), SAC (4-51%), and BESS (2-32%) (Supplemental Table 1.). Performance of the ImPACT subscores (Supplemental Table 2.) showed consistency with previous findings: Verbal Memory (4-34%), Visual Memory (3-30%), Visual Motor Speed (4-34%), and Reaction Time (2-39%). Similar findings were noted on CCAT Processing Speed (2-26), Attention (4-43%), Learning (4-43%), and Working Memory Speed – Speed (4-43%)). There were insufficient data to evaluate the CNS Vital Signs sensitivity when the baseline was captured one year prior.

Among the emerging assessments (Supplemental Table 3), the VOMS showed consistent sensitivity across the subscores with a same season baseline: Smooth Pursuit (41-100%), Horizontal Saccades (37-74%), Vertical Saccades (41-74%), NPC Symptoms (32-74%), NPC(2-48%), Horizontal VOR (34-74%), Vertical VOR (33-76%), and VMS (36-75%). Sensitivity of the K-D (8-36%) and RTclin (8-23%) followed. When change from the baseline assessment captured one year prior was implemented, VOMS performance was similar: Smooth Pursuit (34-100%), Horizontal Saccades (29-78%), Vertical Sacacdes (35-75%), NPC Symptoms (22-74%), NPC(4-51%), Horizontal VOR (28-77%), Vertical VOR (28-74%), and VMS (35-75%). K-D (10-47%) and the RTclin (9-27%) also performed consistently.

<u>3.2 Combined Assessment Performance</u>

Table 2. presents the combined effect of multiple concussion assessments administered at a range of post-injury intervals using CART analyses. The findings show minimal difference between using raw post-concussion scores, the same season baseline difference score, or the prior season baseline difference score. With priority given to the lowest chance of misclassifying a concussed athlete, maximal F1 scores were noted at the Sideline (F1 = 0.639) and Post-event (F1 = 0.658) Evaluations when the SCAT symptom inventory, SAC, and BESS – firm surface only were compared to the same season baseline

evaluation. Administration of those same measures (i.e. SCAT symptoms, SAC and BESS) also provided maximum performance at the 24-72hr interval, but when using the raw score (F1 = 0.580).

To provide clinicians with a clear pathway when baseline evaluations are and are not available during the post-injury evaluation, both the raw score CART analyses and maximally performing baseline comparison CART analyses are presented for each testing interval. In general, SCAT symptoms, both total symptoms and severity, were the primary determinants in correctly identifying concussed and non-concussed athletes. This was followed by administration of the BESS test (firm surface only), and then the SAC in one model.

Insert Table 2 here

*3.2.1 Sideline Evaluation - Raw Scores (BESS-firm surface only, 1:5 ratio)*

CART analyses evaluating the optimal concussion assessment battery and sequence when completed up to 75 minutes following injury was optimized using the raw scores and a 1:5 ratio of concussed to control participant data (Figure 1. - left). The analyses indicated SCAT symptom severity was the most discriminating measure with scores <16 accurately indicating a non-concussed athlete in 4207 out of 4385 cases (96% specificity). When athlete severity scores exceeded 16, a BESS-firm surface only score of 4 or greater indicated correct classification as a concussed athlete in 203 out of 271 cases (75% sensitivity). However, athletes with SCAT symptom severity score greater than 16, and a BESS-firm surface score of 3 or less, were still classified correctly as concussed in 32 out of 36 cases (89% sensitivity) when their SAC total score was less than 24. Among those with SAC scores greater than 24, 25 out of 54 cases (46% sensitivity) were classified as concussed when SCAT symptom severity score exceeded 34. Athletes with SCAT symptom severity score greater than 16, but less than 34, BESS-firm surface only score of 3 or less and SAC total score 24 or more were classified correctly as non-concussed in 144 out of 198 cases (73% specificity).

*3.2.2 Sideline Evaluation - Differences from Same Season Baseline (BESS-firm surface only, 1:5 ratio)*

CART analyses evaluating the optimal concussion assessment battery and sequence when completed up to 75 minutes following injury was optimized when using the change scores from the same season baseline evaluation and a 1:5 ratio of concussed to control participant data (Figure 1. - right). The analyses indicated that the changes in the SCAT symptom severity was the most discriminating measure with score changes less than 10 accurately indicating a non-concussed athlete in 3951 out of 4061 cases (97% specificity). When symptom severity change score were 10 or greater and BESS-firm surface score increased by 3 or more, concussed athletes were correctly categorized in 129 out of 157 cases (82% sensitivity). However, when BESS-firm surface scores changed by 2 or less and symptom severity change score was 16 or greater, 103 out of 190 cases were correctly classified as concussed (54% sensitivity). Athletes with the change in SCAT symptom severity score between 10 and 15 and change in the BESS-firm surface score of 2 or less were classified correctly as non-concussed in 113 out of 150 cases (75% specificity).

Insert Figure 1 here

*3.2.3 Post-event Evaluation - Raw Scores (without computer-based neurocognitive testing or BSI-18, 1:2 ratio)*

CART analyses evaluating the optimal concussion assessment battery and sequence when completed between 75 minutes and 24 hours following injury were optimized when using the raw scores and a 1:2 ratio of concussed to control participant data (Figure 2. - left). The analyses indicated SCAT symptom severity was the most discriminating measure with scores less than 10 accurately indicating a non-concussed athlete in 4017 out of 4269 cases (94% specificity), and SCAT symptom severity scores 10 or greater accurately indicating a concussed athlete in 601 out of 1036 cases (58% sensitivity)

*3.2.4 Post-event Evaluation -Differences from Same Season Baseline (BESS firm surface only and without computer-based neurocognitive testing or BSI-18, 1:2 ratio)*

CART analyses evaluating the optimal concussion assessment battery and sequence when completed between 75 minutes and 24 hours following injury were optimized when using the change scores from the same season baseline evaluation and a 1:2 ratio of concussed to control participant data (Figure 2. – right). SCAT symptom severity was the most discriminating measure with the change scores of 4 or less accurately indicating a non-concussed athlete in 3685 out of 3830 cases (96% specificity). When the changes in the SCAT symptom severity scores were 5 or greater and the change in the total number of reported symptoms was 8 or greater, concussed athletes were correctly classified in 312 out of 416 cases (75% sensitivity). However, when SCAT symptom severity score changes were 5 or greater, but the total number of symptoms were 7 or less, changes in the BESS firm surface only score of 2 or more correctly identified 77 out of 143 concussed athletes (54% sensitivity). Lastly, when SCAT symptom severity scores were 5 or greater, SCAT total symptom score changes were 7 or less BESS firm surface only score change was 1 or less, a SAC total score decrease of 2 or less classified concussed athletes in 30 out of 92 cases (33% sensitivity). Conversely, non-concussed athletes were correctly classified in 262 out of 320 cases (82% specificity) when SCAT symptom severity changes scores exceeded 4, changes in the symptom total were 7 or less, changes in the BESS-firm surface were 1 or less, and the total SAC score decreased by 1 or less.

Insert Figure 2 here

*3.2.5 Clinic Evaluation - Raw Scores (without computer-based neurocognitive testing or BSI-18, 1:2 ratio)*

CART analyses evaluating the optimal concussion assessment battery and sequence when completed between 24 and 72 hours following injury were optimized when using the raw scores and a 1:2 ratio of concussed to control participant data (Figure 3. – left). The analyses indicated SCAT total symptom number was the most discriminating measure with scores of 6 or less indicating a non-concussed athlete in 4038 out of 4591 cases (88% specificity), while scores of 7 or greater accurately classified concussed athlete in 669 out of 1083 cases (62% sensitivity).

*3.2.6 Clinic Evaluation - Differences from Same Season Baseline (without computer-based neurocognitive testing or BSI-18, 1:2 ratio)*

CART analyses evaluating the optimal concussion assessment battery and sequence when completed between 24 and 72 hours following injury were optimized when using the change scores from the same season baseline evaluation and a 1:2 ratio of concussed to control participant data (Figure 3 – right). The analyses indicated that the changes in the SCAT total symptoms was the most discriminating measure with the change scores of 4 or less accurately indicating a non-concussed athlete in 3903 out of 4350 cases (90% specificity), while a total symptom score change of 5 or greater accurately indicating a concussed athlete in 460 out of 736 cases (63% sensitivity).

Insert Figure 3 here

## 4. DISCUSSION

This investigation analyzed data from a large, prospective study on the natural history of concussion to delineate optimal concussion assessments during the acute post-injury interval. Assessment timing was selected based on clinical relevance to represent evaluations on the sideline (0-1.25hrs), post-event (1.25-24hrs), and in a clinic setting (24-72hrs). Consistent with multiple position and consensus statements [9, 16, 8, 7, 1] and previous research [3, 4, 2] our findings indicate that a multifaceted approach to concussion assessment provided the best overall performance in the initial stages of injury, although symptoms alone were the most accurate beyond one day post-injury.

Concussion assessment in the context of competitive sports presents unique challenges. In many instances, the rules of the game do not allow adequate evaluation time, necessitating a rapid evaluation and decision-making process by the sports medicine practitioner. The findings presented herein provide a pathway by which clinicians can administer and interpret findings quickly and efficiently. Importantly, the CART analyses provide a sequential decision-making process requiring the clinician to proceed to the next step only when the assessment results are ambiguous. Indeed, the practical implications of these analyses allows the clinical assessment to be halted once specific criteria indicating concussed or non-concussed status is obtained (Figures 1 to 3). This approach not only optimizes the assessment approach, but also addresses concerns about evaluation duration that continues to be an issue in some sport settings [1]. Importantly, the clinician should always correlate the broader clinical examination with the objective steps presented herein to make a diagnosis [7, 1].

Second to optimizing the concussion assessment tools and sequence was the evaluation of raw post-concussion scores versus change scores calculated from pre-morbid assessments captured one or two seasons prior. F1 scores were highest when change scores were calculated from the same season baseline examination in two of the three testing intervals (Table 2: Sideline and Post-event Evaluations). The third testing interval (Clinic Evaluation) was optimized using raw scores captured from that evaluation. Despite our sample size, insufficient data precluded the calculation of change scores from baseline evaluations captured two seasons prior. Regardless, the diminishing performance noted when baseline data captured one year prior would suggest further accuracy declines in categorizing concussed and non-concussed athletes. This finding suggests that annual baseline testing may optimize post-morbid interpretation, with paramount importance placed on symptom presence and severity, balance evaluated using multiple stances on a firm surface, and neurocognitive screening (i.e., SAC). In the event baseline scores are not available, raw scores can be used to guide the diagnostic process and other work has provided a range of diagnostic certainty under similar circumstances[17]

Fortuitously, the assessments identified collectively as optimal performers by the CART analyses are bundled into the Sport Concussion Assessment Tool, version 5 (SCAT5)[18]. The SCAT5 offers a standardized approach to symptom, balance, and neurocognitive screening while also outlining a protocol for the immediate assessment of the athletes that aids in identifying more severe brain or neck injuries and provides guidance for at-home care. The five-word version of the SAC embedded within the SCAT document was utilized in this investigation, but a 10-word version is available and may produce different findings. Future work should examine this option.

Notable in the analyses was the exclusion of computer-based neurocognitive testing in any of the models, supporting the recommendations by an international panel of concussion experts[1]. The failure of computerized neurocognitive testing to be included in the optimized assessment battery is vexing given its widespread clinical use[19]. Similar to prior results[20], our findings indicate that computerized neurocognitive testing does not uniquely increase the sensitivity of concussion assessment over and above the SCAT symptom assessment and brief screening measures (BESS, SAC) during the acute and early subacute post-injury period. We were unable to evaluate computer-based measure sensitivity in tracking recovery and helping with clinical management beyond 72hrs post-injury, but others have. One investigation reported impairments among 29-39% of concussed athletes reporting as free from concussion-related symptom eight days post-injury and 24-47% impaired at day 15[20]. Those time points are consistent with normal recovery among concussed young adults [1], but are also consistent with false positive findings at the same testing points [20]. Collectively, the routine use of computer-based neuropsychological testing within 72hrs of injury does not appear to be imperative, but complete abandonment is not warranted. There may be instances when specific information can be garnered through computer-based neuropsychological testing both within this post-injury window (i.e., < 72 hours) and beyond (i.e., >72 hours).

A number of emerging baseline and post-injury assessments were completed by concussed athletes, but the lack of universal implementation across all participating schools precluded their use in the primary analyses. Data presented in Supplemental Table 3 outlines their individual sensitivity relative to previously established confidence intervals[15]. As anticipated, a trade-off exists between higher sensitivity and lower specificity and the measures were approximately stable when using change scores from the same season baseline versus calculating change scores from a baseline assessment collected one year prior. As sufficient data were not available to analyze the performance of these assessments to make a clinical use recommendation using the current analytical approach, individual assessment analyses are forthcoming from this group and researchers should continue to evaluate their efficiency in a variety of contexts and refine the instruments based on those findings.

*4.1 Limitations*

As with all investigations, this study is not without limitations. Despite the sample being nationally representative of collegiate athletes, non-varsity level service academy members were not included in these analyses. While some generalization of our findings may be appropriate to that population, the unique demands of the military environment necessitates additional analyses. There is also benefit to replicating this investigation in different cohorts, with a particular emphasis on younger (i.e., high school and adolescent) athletes. Developmental status and trajectories at the younger ages may find different points of interpretation for any number of measures, but particularly where symptoms and motor-control measures are concerned [21, 22]. In addition, a number of pre-morbid conditions (e.g., concussion history) may influence baseline and/or post-injury performance [23]. The breadth of possible co-variates forestalled additional analyses, but should be evaluated by other investigators and taken into consideration by clinicians. Lastly, there is no objective gold standard for concussion diagnosis to which we can compare our findings. Indeed, clinicians rely most on reported symptoms, placing symptom scales among the most important assessment tools. Despite this, our results demonstrate clinical pathways that quantify concussion-related outcomes, but future works should evaluate post-concussion performance on these and other measures among athletes with a diagnosed concussion, but reporting a low symptom burden.

*4.2 Conclusions*

This investigation sought to identify the concussion assessment armamentarium that maximizes sensitivity and specificity among acutely concussed athletes. Our findings indicated that alone, or in combination, a symptom evaluation, BESS scores collected on the firm surface, and SAC total score offered the best overall performance when administered within 72 hours of injury and were compared to pre-morbid performance captured in the same season. The analyses also demarcated a testing sequence with cut points that maximize testing efficiency for the administering clinician and should continue to be used in conjunction with, not as a replacement for, a thorough clinical examination. These findings are consistent with the most recent international guidelines[1] and statements by some medical organizations [9, 16, 8] recommending a multidimensional model of concussion assessment, but diverge from others[7]. Other measures may offer additional insight when administered either within or outside a 72-hour post-injury window or among populations not evaluated here and future work should consider these factors, as well as the uniqueness of other populations.

## COMPLIANCE WITH ETHICAL STANDARDS

**Funding:** This publication was made possible, in part, with support from the Grand Alliance Concussion Assessment, Research, and Education Consortium, funded by the National Collegiate Athletic Association and the Department of Defense. The US Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick, MD 21702-5014, USA is the awarding and administering acquisition office. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Psychological Health and Traumatic Brain Injury Program under Award no. W81XWH-14-2-0151. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense (Defense Health Program funds).

**Conflict of interest**: Steven Broglio, Jaroslaw Harezlak, Barry Katz, Shi Zhao, Thomas McAllister and Michael McCrea received funding from the National Collegiate Athletic Association and the Department of Defense to complete this investigation and cover travel costs related to the study.

**CAPTIONS**

**Table 1:** Mandatory and emerging (i.e., optional) concussion assessments administered at baseline and various post-injury intervals. Sites self-selected one test to administer. *indicates options for computer-based neurocognitive testing.

**TABLE 2:** Sensitivity, specificity, positive predictive value, and F1 score for the classification trees with the largest F1 score among the ratios of concussed to non-concussed equal to 1:2 and 1:5. Classification and Regression Tree (CART) method for a given post-concussion timepoint. Performance was evaluated using the raw post-concussion scores, change score from the same season baseline assessment, and change score from the prior season's baseline score.

**Figure 1:** Left side delineates the CART analysis using raw scores (BESS-firm surface only, 1:5 ratio) at the sideline post-injury evaluation. Difference scores from the same season baseline (BESS-firm surface only, 1:5 ratio) at the 0-1.25 hour post-injury assessment are presented on the right.

**Figure 2**: Left side delineates the CART analysis using raw scores without the BSI-18 or a computer-based neurocognitive assessment (1:2 ratio) at the post-event injury evaluation. Difference scores from the same season baseline without the BSI-18 or a computer-based neurocognitive assessment (BESS-firm stance only, 1:2 ratio) at the same post-injury interval are presented on the right.

**Figure 3:** Left side delineates the CART analysis using raw scores without the BSI-18 or a computer-based neurocognitive assessment (1:2 ratio) at the clinic evaluation. Difference scores from the same season baseline without the BSI or a computer-based neurocognitive assessment (1:2 ratio) at the same post-injury interval are presented on the right.

**Appendix:**

CARE Consortium Investigators are listed alphabetically by institution: Joseph Hazzard (Bloomsburg University), Louise Kelly (California Lutheran University), Darren Campbell, Jonathan Jackson, and Gerald McGinty (US Air Force Academy), Patrick O'Donnell (US Coast Guard Academy), Kenneth Cameron (US Military Academy), Adam Susmarski (US Naval Academy), Josh Goldman and Christopher Giza (University of California–Los Angeles), Thomas Buckley and Thomas Kaminski (University of Delaware), James Clugston (University of Florida), Julianne Schmidt (University of Georgia), Luis Feigenbaum (University of Miami), JT Eckner (University of Michigan), Scott Anderson (University of Oklahoma), Christina Master (University of Pennsylvania), Anthony Kontos (University of Pittsburgh), Sara Chrisman (University of Washington), and Alison Brooks (University of Wisconsin)

**REFERENCES**

1. McCrory P, Meeuwisse W, Dvorak J, Aubry M, Bailes J, Broglio S et al. Consensus statement on concussion in sport-the 5th international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med. 2017. doi:10.1136/bjsports-2017-097699.
2. McCrea M, Barr WB, Guskiewicz KM, Randolph C, Marshall SW, Cantu R et al. Standard regression-based methods for measuring recovery after sport-related concussion. J Int Neuropsychol Soc. 2005;11:58-69.
3. Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. Neurosurgery. 2007;60(6):1050-7; discussion 7-8. doi:10.1227/01.NEU.0000255479.90999.C0.
4. Resch JE, Brown CN, Schmidt J, Macciocchi SN, Blueitt D, Cullum CM et al. The sensitivity and specificity of clinical measures of sport concussion: three tests are better than one. BMJ open sport & exercise medicine. 2016;2(1):e000012. doi:10.1136/bmjsem-2015-000012.
5. Resch J, Driscoll A, McCaffrey N, Brown C, Ferrara MS, Macciocchi S et al. ImPact test-retest reliability: reliably unreliable? J Athl Train. 2013;48(4):506-11. doi:10.4085/1062-6050-48.3.09.
6. Kerr ZY, Register-Mihalik JK, Kroshus E, Baugh CM, Marshall SW. Motivations Associated With Nondisclosure of Self-Reported Concussions in Former Collegiate Athletes. The American journal of sports medicine. 2015. doi:10.1177/0363546515612082.
7. Broglio SP, Cantu RC, Gioia GA, Guskiewicz KM, Kutcher J, Palm M et al. National Athletic Trainers' Association position statement: management of sport concussion. J Athl Train. 2014;49(2):245-65. doi:10.4085/1062-6050-49.1.07.
8. Giza CC, Kutcher JS, Ashwal S, Barth J, Getchius TS, Gioia GA et al. Summary of evidence-based guideline update: Evaluation and management of concussion in sports: Report of the Guideline Development Subcommittee of the American Academy of Neurology. Neurology. 2013;80(24):2250-7.
9. Harmon KG, Clugston JR, Dec K, Hainline B, Herring S, Kane SF et al. American Medical Society for Sports Medicine position statement on concussion in sport. Br J Sports Med. 2019;53(4):213-25. doi:10.1136/bjsports-2018-100338.
10. Chin EY, Nelson LD, Barr WB, McCrory P, McCrea MA. Reliability and Validity of the Sport Concussion Assessment Tool-3 (SCAT3) in High School and Collegiate Athletes. The American journal of sports medicine. 2016;44(9):2276-85. doi:10.1177/0363546516648141.
11. Zimmer A, Piecora K, Schuster D, Webbe F. Sport and team differences on baseline measures of sport-related concussion. J Athl Train. 2013;48(5):659-67. doi:10.4085/1062-6050-48.5.06.
12. Hunt TN, Ferrara MS. Age-related differences in neuropsychological testing among high school athletes. J Athl Train. 2009;44(4):405-9.
13. Broglio SP, McCrea M, McAllister T, Harezlak J, Katz B, Hack D et al. A National Study on the Effects of Concussion in Collegiate Athletes and US Military Service Academy Members: The NCAA-DoD Concussion Assessment, Research and Education (CARE) Consortium Structure and Methods. Sports Med. 2017;47(7):1437-51. doi:10.1007/s40279-017-0707-1.
14. Carney N, Ghajar J, Jagoda A, Bedrick S, Davis-O'Reilly C, du Coudray H et al. Concussion guidelines step 1: systematic review of prevalent indicators. Neurosurgery. 2014;75 Suppl 1:S3-15. doi:10.1227/neu.0000000000000433.
15. Broglio SP, Katz BP, Zhao S, McCrea M, McAllister T, Investigators CC. Test-Retest Reliability and Interpretation of Common Concussion Assessment Tools: Findings from the NCAA-DoD CARE Consortium. Sports Med. 2018;48(5):1255-68. doi:10.1007/s40279-017-0813-0.
16. Herring SA, Cantu RC, Guskiewicz KM, Putukian M, Kibler WB, Bergfeld JA et al. Concussion (mild traumatic brain injury) and the team physician: a consensus statement--2011 update. Medicine and Science in Sports and Exercise. 2011;43(12):2412-22.

17. Garcia GP, Broglio SP, Lavieri MS, McCrea M, McAllister T, Investigators CC. Quantifying the Value of Multidimensional Assessment Models for Acute Concussion: An Analysis of Data from the NCAA-DoD Care Consortium. Sports Med. 2018;48(7):1739-49. doi:10.1007/s40279-018-0880-x.
18. Echemendia RJ, Meeuwisse W, McCrory P, Davis GA, Putukian M, Leddy J et al. The Sport Concussion Assessment Tool 5th Edition (SCAT5). Br J Sports Med. 2017. doi:10.1136/bjsports-2017-097506.
19. Resch JE, McCrea MA, Cullum CM. Computerized neurocognitive testing in the management of sport-related concussion: an update. Neuropsychol Rev. 2013;23(4):335-49. doi:10.1007/s11065-013-9242-5.
20. Nelson LD, LaRoche AA, Pfaller AY, Lerner EB, Hammeke TA, Randolph C et al. Prospective, Head-to-Head Study of Three Computerized Neurocognitive Assessment Tools (CNTs): Reliability and Validity for the Assessment of Sport-Related Concussion. J Int Neuropsychol Soc. 2016;22(1):24-37. doi:10.1017/s1355617715001101.
21. Williams RM, Puetz TW, Giza CC, Broglio SP. Concussion recovery time among high school and collegiate athletes: a systematic review and meta-analysis. Sports Med. 2015;45(6):893-903. doi:10.1007/s40279-015-0325-8.
22. Davis GA, Anderson V, Babl FE, Gioia GA, Giza CC, Meehan W et al. What is the difference in concussion management in children as compared with adults? A systematic review. Br J Sports Med. 2017;51(12):949-57. doi:10.1136/bjsports-2016-097415.
23. Brooks BL, Mannix R, Maxwell B, Zafonte R, Berkner PD, Iverson GL. Multiple Past Concussions in High School Football Players: Are There Differences in Cognitive Functioning and Symptom Reporting? The American journal of sports medicine. 2016;44(12):3243-51. doi:10.1177/0363546516655095.